JURNAL NUANSA INFORMATIKA
Volume 16 Nomor 1, Januari 2022
p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit

Mohammad Rizal Givari[1], Mochamad Riszky Sulaeman[2], Yuyun Umaidah[3].
[1]Universitas Singaperbangsa Karawang
E-mail:[1]mohammad.rizal18257@student.unsika.ac.id,[2]mochamad.riszky18165@student.unsika.ac.id,[3]yuyun.umaidah@staff.unska.ac.id

Abstrak

Kredit merupakan salah satu opsi untuk mencari pendanaan pada kebanyakan kegiatan ekonomi. Permintaan kredit saat ini sudah berkembang dengan sangat pesat, sejalan dengan kebutuhan finansial di kalangan masyarakat yang semakin meningkat khususnya di negara berkembang seperti Indonesia. Analisis kredit perlu dilakukan untuk mencapai pemberian kredit yang tepat dan aman. Analisis kredit merupakan pengamatan guna melihat kelayakan dari sebuah permasalahan kredit. Dari analisis tersebut, akan diketahui kelayakan penerima kredit. Pada penelitian ini menggunakan metodologi CRISP-DM yang teridir dari 6 tahap, yaitu Bussines Understanding, Data Understanding, Data preparation, Modelling Evaluation, dan Deployment dengan menerapkan metode klasifikasi dengan membandingkan algoritma SVM, Random Forest, dan XGBoost. Pada penelitian ini menggunakan dataset yang bersifat open source yang diperoleh dari Kaggle. Hasil penelitian dengan menggunakan algoritma SVM, random forest, dan XGBoost mendapatkan nilai accuracy, recall, precision tertinggi pada model XGBoost dengan nilai accuracy sebesar 82%, recall 70%, dan precision 92%.

***Kata Kunci**: Kredit, SVM, Random forest, XGBoost.*

Abstract

Having a credit is an option for seeking funding for most economic activities recently. The demand for having a credit is currently growing very rapidly, in line with the increasing financial needs of the community, especially in developing countries such as Indonesia. Credit analysis is needed to be carried out to achieve fit and proper credit matters. Credit analysis is an observation to see the feasibility of a credit problem. From this analysis, the creditworthiness of the recipient will be known. This study uses the CRISP-DM methodology which consists of 6 stages, namely Bussines Understanding, Data Understanding, Data preparation, Modeling Evaluation, and Deployment by applying the classification method by comparing the SVM, Random Forest, and XGBoost algorithms. This research uses an open source dataset obtained from Kaggle. The results of the research using the SVM, random forest, and XGBoost algorithms get the highest accuracy, recall, precision values in the XGBoost model with 82% accuracy, 70% recall, and 92% precision.

***Keywords:** Credits, SVM, Random forest, XGBoost.*

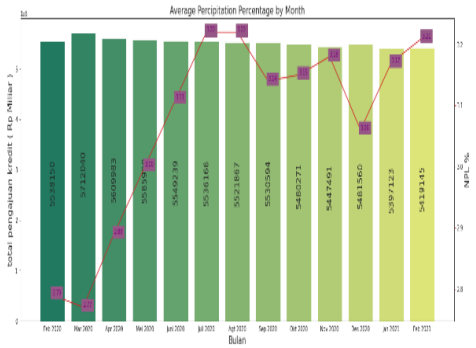
1. PENDAHULUAN

Knowledge Discovery in Database (KDD) atau yang lebih dikenal dengan data mining merupakan suatu penyelesaian masalah dengan melakukan analisis terhadap data yang disajikan dalam database. Selain itu data mining juga digunakan untuk mengetahui pola data, dimana setiap pola memiliki karakteristik masing-masing yang dapat memberikan informasi penting dari data tersebut. Data mining dapat diartikan sebagai berbagai macam cabang ilmu pengetahuan yang menjadi satu, terdiri atas sistem basis data, statistika, machine learning, visualization, dan informasi pengetahuan. Data mining telah berhasil diterapkan diberbagai bidang ilmu seperti ekonomi, bioinformatika, genetika, kedokteran, pendidikan dan lain sebagainya. Salah satu contoh penerapan data mining pada bidang ekonomi ialah untuk klasifikasi nasabah bank dalam menentukan keputusan pemberian kredit.

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

Permintaan kredit saat ini sudah berkembang dengan sangat pesat, sejalan dengan kebutuhan finansial di kalangan masyarakat yang semakin meningkat khususnya di negara berkembang seperti Indonesia. Ini diperkuat dengan survei yang dilakukan oleh Bank Indonesia pada triwulan II 2021 yang menyebutkan bahwa terdapat kenaikan permintaan pengajuan kredit sebesar 53.9%. Berdasarkan UU No.10 Tahun 1998 yang menjelaskan bahwa kredit adalah penyediaan uang, tagihan dan yang lainnya atas kesepakatan antara bank dengan peminjam dengan jangka waktu dan bunga dalam melunasi hutangnya. Dalam dunia perbankan, penga-daan kredit terhadap nasabah memung-kinkan adanya resiko tinggi, maka dari itu pengelolaannya dianggap sebagai tugas penting.

Dalam realisasinya, analisis kredit yang kurang hati-hati dalam proses tersebut, mengakibatkan kredit bermasalah. Kesalahan analisa kredit dapat menyebabkan risiko kredit [1]. Sesuai dengan Otoritas Jasa Keuangan (2016) yang mengemukakan bahwa Risiko kredit adalah kegagalan pihak peminjam untuk dapat memenuhi kewajibannya terhadap bank, termasuk risiko kredit yang diakibatkan gagalnya debitur, risiko konsentrasi kredit, *counterparty credit risk*, dan *settlement risk*. Dari penelitian yang dilakukan Otoritas Jasa Keuangan (OJK) yang dilampirkan pada situs resmi mereka, menyatakan bahwa terdapat kenaikan NPL (Non Performing Loan) pada rentang Feb 2020 sampai dengan Feb 2021, seperti yang dilampirkan pada grafik dibawah ini.



Gambar 1.Average Precipitation Percentage by Month

NPL itu sendiri merupakan pinjaman yang bermasalah dimana peminjam tidak dapat membayarkan sebuah pinjaman sesuai jadwal yang telah disepakati, sehingga berdampak pada sisi keuangan/financial bank yang memberikan pinjaman menjadi bermasalah. Analisis kredit perlu dilakukan untuk mencapai pemberian kredit yang tepat dan aman. Analisis kredit merupakan pengamatan guna melihat kelayakan dari sebuah permasalahan kredit. Dari analisis tersebut, akan diketahui kelayakan penerima kredit.

Untuk menentukan kelayakan kredit dapat diaplikasikan salah satu Teknik Klasifikasi pada Data Mining. Larose, D. T., & Larose, C. D.[2] menulis bahwa data mining merupakan serangkaian proses dengan memilih data menggunakan teknologi, teknik statistika dan matematika untuk menemukan hubungan, pola, ataupun tren yang mempunyai makna. Diharapkan dengan adanya penggunaan data mining membantu dalam memprediksi resiko kelayakan kredit. Kita juga dapat memanfaatkan model hasil klasifikasi untuk memprediksi tren masa depan [3].

Terdapat penelitian terdahulu yang membahas mengenai risiko kredit. Pada tahun 2016 penelitian dilakukan oleh Mittal, Gupta, & Sangaiah [4] menggunakan tiga metode, yaitu SVM, Neural Network, dan Naive Bayes untuk menentukan algoritma yang paling optimal untuk memprediksi risiko kredit. Penelitian ini bertujuan untuk penentu keputusan yang tepat dalam membantu mengurangi kerugian pada manajemen bank. Pada Penelitian ini SVM menggunakan polynomial kernel, Neural Network dengan iterasi 500, dan Naive Bayes dengan sekali proses. Penelitian ini menghasilkan tingkat akurasi tertinggi oleh SVM sebesar 92%, Naive Bayes 87%, dan Neural Network sebesar 85%.

Penelitian kedua yang dilakukan oleh Chakraborti [5]. Penelitian ini membandingkan beberapa algoritma untuk

Terakreditasi SINTA 5

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

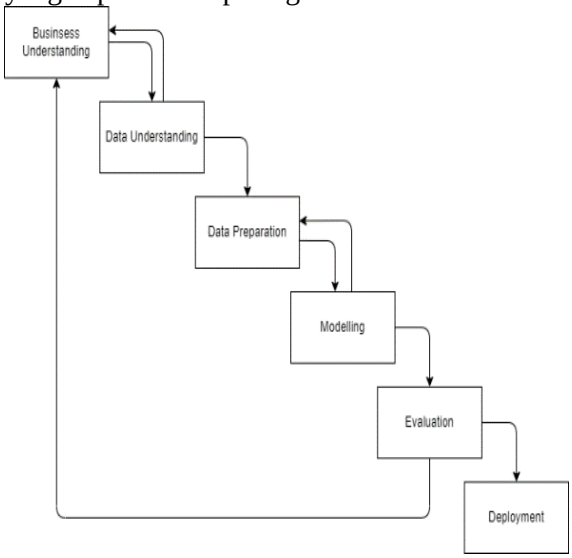
menentukan algoritma yang paling akurat untuk memprediksi gaji karyawan. Metode yang digunakan pada penelitian ini adalah Naive Bayes, SVM, dan Neural Network menghasilkan akurasi tertinggi pada SVM dengan 84,9%, Naive Bayes menghasilkan 83,4% dan Neural Network dengan hasil 82,89%.

Pada penelitian yang dilakukan Bonggo & Wasono dilakukan perbandingan nilai akurasi dari dua algoritma yaitu Random Forest dan Naibe Bayes. Dari penelitian ini mendapatkan hasil akurasi yaitu 98,16% [6].

Pada tahun 2019, penelitian dilakukan oleh Hanif,I. Pada penelitian ini bertujuan untuk membuktikan apakah metode XGBost dapat melakukan prediksi lebih baik dengan Regresi Logistik. Data yang digunakan berasal dari perusahan Telkom Indonesia [7].

2. METODE PENELITIAN

Pada penelitian ini menggunakan metodologi CRISP-DM dengan menerapkan metode klasifikasi dengan membandingkan algoritma SVM, Random Forest, dan XGBoost. CRISP-DM (*Cross Industry Standard Process Model for Data mining*) adalah sebuah metode yang banyak digunakan ahli dengan model proses pengembangan data dalam pemecahan masalah [8]. CRISP-DM mempunyai 6 fase/tahapan pada proyek data mining [9]. Penelitian ini dilakukan dengan 6 tahapan yang dapat dilihat pada gambar 2.1 berikut.



Gambar 2.Tahapan Penelitian

2.1 Business Understanding

Kredit kepada masyarakat bukanlah masalah yang aneh, kredit merupakan salah satu opsi untuk mencari pendanaan pada kebanyakan kegiatan ekonomi. Sebelum memberikan kredit kepada nasabah, harus ada proses untuk mengidentifikasi dan memperkirakan secara akurat dan lengkap semua aspek kredit yang dapat membantu proses pemberian pinjaman, untuk mencegah timbulnya risiko kredit. Karenanya diperlukan suatu kegiatan untuk meminimalisir terjadinya masalah pada pemberian kredit , salah satunya dapat diatasi dengan melakukan identifikasi dan forecasting nasabah dengan baik sebelum memberikan pinjaman dengan memperhatikan fokus pada data historis pemberian pinjaman. Oleh karena itu, klasifikasi risiko kredit di sektor perbankan memiliki peran penting. Jika terjadi kesalahan dalam klasifikasi risiko kredit, salah satu dampaknya adalah kredit macet, dan kredit buruk atau kredit yang macet dapat menyebabkan masalah finansial pada bank.

2.2 Data Understanding

Pengumpulan dan memahami data dilakukan pada tahap ini untuk mendapatkan *insight* dan kualitas dari data yang peneliti miliki. Untuk *dataset* pada penelitian ini menggunakan *dataset* yang bersifat *open source* yang diperoleh dari *Kaggle.*

Judul	*Credit Card Dataset*
Jumlah Fitur (Kolom)	20
Bentuk Data	CSV

Tabel 1. Data

2.3 Data Preparation

Tahap ketiga merupakan tahap yang paling penting pada metode Crisp-DM ini, yaitu data preparation. pada

Terakreditasi SINTA 5

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

tahap ini kita akan menyiapkan data yang sebelumnya masih mentah menjadi data yang siap digunakan. Terdapat beberapa proses pada tahap ini, yaitu :

a. *Data Selection* : proses ini merupakan proses untuk mengidentifikasi dan memilih fitur atau kolom mana saja yang akan digunakan yang akan berdampak baik untuk hasil klasifikasi kita serta menghilangkan fitur atau kolom yang tidak ada korelasi dengan tujuan pada penelitian ini [10].
b. *Data Preprocessing :* setelah kita memilih fitur atau kolom pada proses sebelumnya langkah selanjutnya adalah proses data *preprocessing*. Pada proses ini kita akan mengidentifikasi *missing value, outliers* yang ada pada data setelah itu kita lakukan *cleaning* atau pembersihan data dari *missing value* dan *outliers.*
c. Data *Transforming* : sebuah model data mining akan bekerja dengan baik jika data - data yang kita punya berbentuk angka, untuk itu diperlukan transformasi data jika terdapat nilai selain numerik pada data kita. Dan juga data numerik dengan skala yang berbeda jauh diantara fitur juga akan mempengaruhi performa model, untuk mengatasi masalah tersebut kita dapat melakukan proses transformasi data yang biasa disebut dengan scaling. *Scaling* merupakan proses dimana menskalakan data pada tiap fitur agar tidak terlalu berbeda jauh, terdapat beberapa teknik scaling, yaitu normalisasi, min-max scaling.

2.4 Modeling

Tahap selanjutnya adalah proses *modeling*, modeling ini merupakan proses pemilihan model yang tepat untuk menyelesaikan masalah pada penelitian ini. *Model* adalah sebuah algoritma yang mempresentasikan perhitungan matematika dalam bahasa pemrograman. Model yang akan digunakan antara lain *Support Vector machine (SVM), Random Forest, XGBoost.*

1. *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah sebuah algoritma klasik yang bertujuan agar dapat menyelesaikan masalah klasifikasi [11]. Algoritma ini merupakan salah satu metode supervised learning. Dibandingkan dengan teknik klasifikasi lainnya, SVM ini memiliki konsep matematika yang lebih matang, sehingga bisa mengatasi masalah klasifikasi linear dan non linear [12].

SVM memiliki prinsip bisa melakukan klasifikasi kedalam dua kelompok dengan menentukan hyperplane yang tepat. Rumus perhitungan SVM adalah sebagai berikut:

1. Titik data :
   $x_i = \{ x_1, x_2, \ldots\ldots\ldots\ldots, x_n \} \in R^n$
2. Kelas data : $y_i \in \{-1, +1\}$
3. Pasangan data dan kelas :$\{(x1, y1)\}_{i=1}^{N}$
4. Maksimalkan fungsi:
   $Ld = \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$
   $syarat : 0 \leq \alpha_i \leq C \; dan \; \sum_{i=1}^{N} \alpha_i y_i$
5. Menghitung nilai w dan b :
6. $w = \sum_{i=1}^{N} \alpha_i y_i x_i \quad b = -\frac{1}{2} w.x^{+} + w.x^{-}$
7. Fungsi Keputusan klasifikasi sign (f(x)):
   $f(x) = w.x + b$ atau $f(x) = \sum_{i=1}^{m} \alpha_i y_i K(x, x_i) + b$

Keterangan :
N : Jumlah data
N : Dimensi data atau banyaknya fitur
Ld : Dualitas Lagrange Multiplier
α_i : Nilai bobot setiap titik data
C : Nilai konstanta
m : Jumlah support vector/titik data yang memiliki $\alpha_i > 0$
$K(x,x_i)$: fungsi kernel

2. *Random Forest*

Random Forest merupakan algoritma yang dibangun dari beberapa *decision tree*, pada dasarnya algoritma ini termasuk *supervised learning* . *Random forest* adalah satu jenis metode yang dapat digunakan untuk klasifikasi dan regresi. Metode *Random Forest* atau *random ensemble* yaitu kombinasi pohon keputusan [13].
Terdapat tiga aspek penting yaitu:

Terakreditasi SINTA 5

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

(1) Membangun pohon prediksi dengan dilakukannya bootstrap sampling
(2) Setiap pohon keputusan memprediksi menggunakan prediktor acak
(3) *Random forest* akan melakukan prediksi dengan kombinasi hasil dari setiap pohon dengan majority vote untuk klasifikasi atau rata-rata untuk regresi [14].

Perhitungan nilai *entropy* memakai rumus pada persamaan 1 dan *information gain* pada persamaan 2 [15].

$$Entropy\ (Y) = -\sum_i \quad \rho\ (c\ |Y)\ log_2\ \rho(c|Y) \tag{1}$$

Y merupakan himpunan kasus dan (p(c|Y) adalah proporsi nilai Y pada kelas c.

$$Information\ Gain\ (Y, a) = Entropy\ (Y) - \sum_{v \in Values(a)} \frac{Y_v}{Y_a}\ Entropy(Y_v) \tag{2}$$

Values(a) adalah semua nilai yang mungkin pada himpunan kasus a. Y_v merupakan subkelas dari Y dengan kelas v yang berkaitan dengan kelas a. Y_a merupakan semua nilai yang sama dengan a.

3. *XGBoost*

Metode XGBoost adalah sebuah metode yang telah ditemukan oleh Friedman. Metode ini adalah pengembangan dari algoritma GBDT (*Gradient Boosting Decision Tree*) [16]**.** Xgboost merupakan salah satu pustaka machine learning yang dapat difungsikan untuk memprediksi atau mengklasifikasikan berbasis pohon keputusan [17]. Algoritma ini memungkinan melakukan optimasi 10 kali lebih cepat dibandingkan dengan GBM lainnya [18]. Sedangkan untuk menghasilkan nilai akurasi hasil klasifikasi tergantung dengan parameter yang digunakan.

XGBoost dan *random forest* merupakan algoritma yang tersusun dari beberapa *decision tree*. Berbeda dengan *random forest* yang menggunakan *bagging*, *XGBoost* ini menggunakan teknik *boosting* dalam penyusunan algoritmanya [19].

2.5 Evaluation

Pada fase ini dilakukan analisis terhadap hasil dari proses pembelajaran data. Fase ini yaitu proses interpretasi hasil pemodelan data mining yang digunakan. Evaluasi Model dilakukan dengan melihat dari *Confusion Matrix* dan juga *ROC Curve (Receiver Operating Characteristic).*

a. Confusion Matrix

Confusion matrix adalah sebuah metode yang mempresentasikan hasil melalui tabel matriks untuk perhitungan akurasi, *recall*, *precision* dan *error*.

		True Class	
		Positive	*Negative*
Predicted Class	*Positive*	*True Positive (TP)*	*False Negative (FN)*
	Negative	*False Negative (FN)*	*True Negative (TN)*

Tabel 2.*Confusion Matrix*

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

b. ROC

ROC adalah sebuah metode menggambarkan dan klasifikasi kategori pada sebuah model statistik. Selain itu juga dapat menentukan *threshold* dari sebuah model. ROC berada pada tabel *confusion matrix* antara *False Positive* dalam garis horizontal terhadap *True Positif* yang akan ditampilkan dalam garis vertikal [20].

2.6 Deployment

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

Deployment merupakan penyusunan laporan atau presentasi dari *modelling* serta *evaluation* pada data Mining. Hasil ini dapat digunakan untuk menentukan keputusan yang tepat dalam mengidentifikasi dan memprediksi yang baik dan seksama terhadap semua aspek perkreditan yang dapat menunjang proses pemberian kredit, guna mencegah timbulnya masalah risiko kredit.

4. HASIL DAN PEMBAHASAAN

Dataset pada pelatihan ini dibagi menjadi dua, yaitu data train dan data test dengan skenario 80% data training dan 20% data test. Dan sebelum membagi/*splitting* menjadi dua, dataset terlebih dahulu dilakukan teknik SMOTE, karena data tersebut imbalance. Pada penelitian ini model yang akan dibandingkan adalah SVM, Random Forest, dan XGBoost. Pada Tabel 3 merupakan rangkuman dari hasil evaluasi dari komparasi tiga algoritma tersebut.

Model	Performa				
	Accuracy	Precision	Recall	F1-score	ROC
SVM	0.69	0.72	0.59	0.65	0.758
Random Forest	0.71	0.72	0.69	0.70	0.752
XGBoost	0.82	0.92	0.70	0.80	0.84

Tabel 3.Hasil

Pada tabel 3 menunjukan bahwa model XGBoost memberikan performa yang lebih paling baik dari ketiganya, model ini mendapatkan nilai 0.82 pada *accuracy*, 0.92 pada *precision*, 0.70 pada *recall*, 80 pada *f1-score*, dan 0.84 pada ROC, berikut penjelasan secara detailnya.

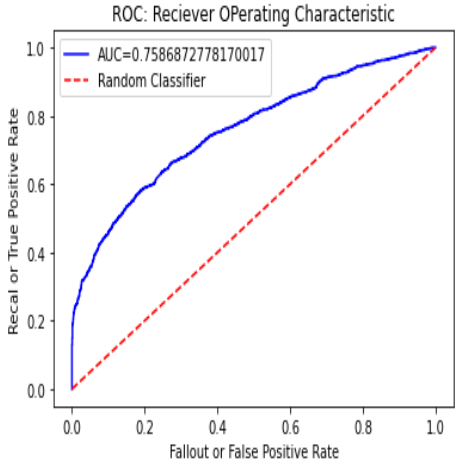
a. SVM

Pada penelitian ini model SVM dilatih dengan menggunakan *hyperparameter* sebagai berikut :

kernel	rbf
C	1000
gamma	0.1

Tabel 4.*hyperparameter* Model SVM

Dengan hyperparameter di atas model SVM yang dilatih menghasilkan performa yang tidak begitu baik. Model SVM ini hanya mendapatkan nilai *accuracy* sebesar 0.69, *f1-sco*re sebesar 0.65, *recall* sebesar 0.59, *precision* sebesar 0.72.



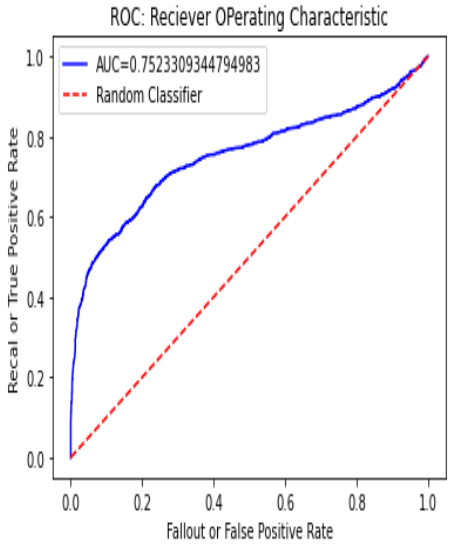
Gambar 3. ROC SVM

Pada gambar diatas menunjukan grafik ROC, plot biru menunjukan akurasi model, dan garis putus - putus merah menunjukan batas wajar performa model. Semakin garis biru mendekati garis merah ataupun berada di bawah garis merah, itu menunjukan bahwa model yang dilatih memiliki performa yang buruk. Pada gambar tersebut menunjukan bahwa garis biru berada tidak berada terlalu jauh dari garis merah. Dengan *score* AUC hanya sebesar 0.75.

b. Random Forest

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

Pada penelitian model *random forest* dilatih dengan n_estimator sebanyak 200, karena random forest terdiri dari beberapa *decision tree* maka n_estimator ini menunjukan jumlah *decision tree* yang digunakan pada model *random forest*. Dan hasil penelitian ini *random forest* memiliki performa yang tidak begitu baik tapi lebih baik dibandingkan model SVM. Model *random forest* ini mendapatkan accuracy sebesar 0.71, *f1-score* sebesar 0.70, *recall* sebesar 0.69, *precision* sebesar 0.72.



Gambar 4.ROC *Random Forest*

Pada grafik diatas random forest mendapatkan nilai ROC yang tidak berbeda jauh dibanding SVM, model ini mendapatkan nilai AUC sebesar 0.75.

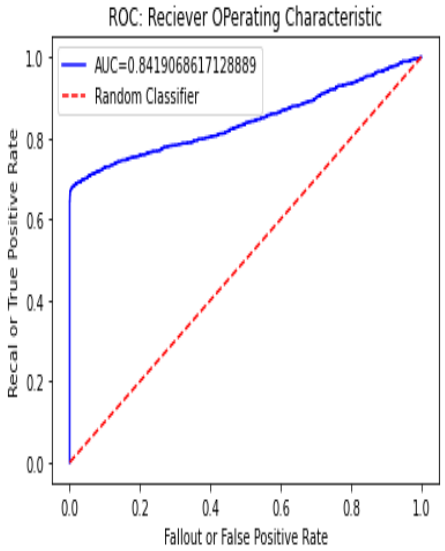
c. XGBoost

Pada penelitian ini model XGBoost dilatih dengan *hyperparameter* sebagai berikut

best score	0.5
booster	gbtree
Learning rate	0.3
n_estimator	100
n_jobs	12

Tabel 5. Hyperparameter XGBoost

Dengan *hyperparameter* diatas model XGBoost berhasil mendapatkan performa yang jauh lebih baik dibandingkan dua model sebelumnya. Model ini mendapatkan nilai *accuracy* sebesar 0.82, *f1-score* sebesar 0.80, *recall* sebesar 0.70, dan *precision* sebesar 0.92.



Gambar 5.ROC XGBoost

Bahkan pada grafik ROC pun model ini memiliki performa yang baik, terlihat bahwa nilai AUC model ini sebesar 0.84, yang jauh melebihi dua model sebelumnya.

5. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan untuk memprediksi kelayakan kredit, dapat diambil kesimpulan sebagai beikut:

a. Algoritma SVM, *Random forest* dan XGBoost mendapatkan nilai *accuracy*, *recall*, *precision* dengan nilai diatas 50%.
b. Hasil penelitian dengan menggunakan algoritma SVM, random forest, dan XGBoost mendapatkan nilai *accuracy*, *recall*, *precision* tertinggi pada model XGBoost dengan nilai *accuracy* sebesar 82%, *recall* 70%, dan *precision* 92%.

6. SARAN

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

Sebaiknya dikombinasikan lebih banyak lagi dalam pengguanaan metode dalam anaslias data dan juga penelitian ini dapat dikembangkan dengan menggunakan algoritma lain dalam data mining.

DAFTAR PUSTAKA

[1] Menarianti, I. (2015). Klasifikasi Data Mining Dalam Menentukan Pemberian Kredit Bagi Nasabah Koperasi. *Jurnal Ilmiah Teknosains*, 26-45.

[2] Larose, D. T., & Larose, C. D. (2014). *Discovering Knowledge In Data An Introduction To Data Mining.*

[3] Y. Pristyanto, "Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritma Klasifikasi Pada Imbalanced Dataset," J. *TEKNOINFO*, 13, no. 1, pp. 11–16, 2019, doi: 10.33365/jti.

[4] Mittal, L., Gupta, T., & Sangaiah, A. K. (2016). PREDICTION OF CREDIT RISK EVALUATION USING NAIVE BAYES,. *The IIOAB Journal*, 33-42.

[5] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," Pattern Recognit., vol. 40, no. 12, pp. 3358–3378, 2007, doi: 10.1016/j.patcog.2007.04.009.

[6] Bawono, B., & Wasono, R. (2019 (3)). PERBANDINGAN METODE RANDOM FOREST DAN NAÏVE BAYES UNTUK KLASIFIKASI DEBITUR BERDASARKAN KUALITAS KREDIT. *Seminar Nasional Edusaintek*, 343-348.

[7] Hanif, I. (2019). Implementing Extreme Gradient Boosting (XGBoost) Classifier to Improve Customer Churn Prediction. International Conference on Statistics and Analytics.

[8] Astuti, D., Iskandar, A. R., & Febrianti, A. (2019). Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA), 1(2)*, 060-072.

[9] Feblian, D., & Daihan, D. U. (2016). Implementasi Model CRISP-DM untuk Menentukan Sales Pipeline pada PT X. *Jurnal Teknik Industri, 1(1)*, 1-12.

[10] Fahmi, R. N., Jajuli, M., & Sulistiyowati, N. (2021). Analisis Pemetaan Tingkat Kriminalitas di Kabupaten Karawang menggunakan Algoritma K-Means. *INTECOMS: Journal of Information Technology and Computer Science*, 67 - 79.

[11] Gaye, B., & Zhang, D. W. (2021). Improvement of Support Vector Machine Algorithm in Big Data Background. *Mathematical Problems in Engineering*.

[12] Prajarini, D. (2016). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Kulit. *INFORMAL: Informatics Journal*, 1-5.

[13] Umar, R., & Riadi, I. P. (2020). Perbandingan Metode SVM, RFdan SGDuntuk Penentuan Model Klasifikasi Kinerja Programmerpada Aktivitas Media Sosial. *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 329-325.

[14] Sadewo, M. G., Windarto, A. P., & Hartama, D. (2017). Penerapan Datamining Pada Populasi Daging Ayam RAS Pedaging di Indonesia Berdasarkan Provinsi Menggunakan K-Means Clustering. *InfoTekJar (Jurnal Nasional Informatika dan Teknologi Jaringan)*, 60-67.

[15] Nugroho, Sulistyo, Yusuf. Emiliyawati, Nova. 2017. Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest. *Jurnal Teknik Elektro.9(1).*

[16] Friedman, J. H. (2001). Greedy Function Approximation: A Gradient

p-ISSN : 1858-3911, e-ISSN : 2614-5405
https://journal.uniku.ac.id/index.php/ilkom

Boosting Machine. *The Annals of Statistics*, 1189-1232.

[17] Jiang, Y., Tong, G., Yin, H., & Xiong, N. (2019). A Pedestrian Detection Method Based on Genetic Algorithm for Optimize XGBoost Training Parameters. *IEEE Access*, 118310 - 118321.

[18] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

[19] M. Syukron, R. Santoso, & T. Widiharih,(2020). "Perbandingan Metode Smote Random Forest Dan Smote Xgboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C Pada Imbalance Class Data", *Jurnal Gaussian*,9, 227- 236.

[20] Mohammadi, N., & Zangeneh, M. (2016). Customer Credit Risk Assessment using Artificial Neural Networks. *International Journal of Information Technology and Computer Science*, 58-66.